

第三章 信息技术教学评价

第一节 信息技术评价基础

一、基本概念的界定

(一)测量、测验、考试、评价

1. 测量

早在 1951 年,史迪文森在其《实验心理学手册》中提出“广义而言,测量是根据法则给事物赋予数量。”也就是说,用一定规则给事物属性指派数字或符号的过程即测量。这是迄今为止公认的测量定义。根据这一定义,测量通常包含三个要素:

(1) 事物及属性

事物及属性即测量的对象或目标。教育和心理测量往往是通过测量个体的外显行为或外在表现特征,来探知隐含于所外显行为之中的个体潜在特质及其水平,比如说某次表现性评价测量是学生在某个信息问题解决任务上的表现。但测量者真正想测查的是学生的信息素养。因此教育与心理测量中的事物属性是指个体的心理特质,是间接测量的结果。

(2) 法则

法则即测量所依据的规则和方法,法则是测量的关键。法则的选择直接决定测量是否可靠,好的法则可保证得到可靠的测量,差的法规则往往会得到不可靠

的甚至是错误的结果。法则的好坏取决于它是否符合被测事物的属性和规律,以及是否易于制定和便于操作。由于人类的心理特质抽象易变,其测量规则的制定和应用相对困难得多。

(3) 数字或符号

数字是代表某一事物或事物某一属性的量。数字本身只是一种符号,只有当我们赋予它意义时,它才变成量化的数。数具有自然数系统的某些特点,包括:区分性,例如“相等”“不相等”等;等级性,例如, $1 < 2 < 3$;等距性, $2 - 1 = 3 - 2$;可加性, $1 + 1 = 2$,等等。由于数字本身具有这些逻辑运算的特征,所以通过测量所得的数,不仅可以表示事物属性的类别、大小、多少,而且还可以在一定的条件下由数的运算而对事物的属性进行推测。

任何领域里的测量都应包含上述三元素,教育评价也是利用测量原理对教育领域内的事物或现象给予数字化的描述,比如学生的学业成就、学习能力水平等。可以说,测验则是教育评价的主要量化工具。

2. 测验

对测验有多种多样的定义,阿娜斯塔西 1982 年在其《心理测验》中所下的定义比较完整,至今仍为大多数教育和心理测验学家所接受,而且适于对高等院校教学中使用的测验进行解释:“测验实质上是对行为样本的客观的和标准化的测量。”根据这一定义,测量与测验是包含关系,测验是对测量对象、测验程序、测验的客观性和规范性做了明确界定的测量,也就是说,一个测量要成其为测验,必须包含三个基本界定因素:

(1) 行为样本

高等院校教学中采用测验是希望了解学生对课程规定的学习目标的掌握情况,但高等院校学习目标是多层次的,而且涉及的内容十分广泛,测验不可能涵盖学生在某一阶段学习内容和学习目标中所有可能的行为表现,因此测验只能选取一组有代表性的内容和目标,以学生在这些内容和目标上的行为作为行为样本来

考查学生在相应行为领域的行为特征。当学生的行为样本很恰当地代表测验所要测的全部内容时,该测验就会提供了有用的预测信息,否则,就说明构成测验的行为样本代表性不强。

(2) 标准化

标准化是指测验在编制、施测、计分及分数解释方面应依据一套系统规范的程序,保证测验有统一的标准,并使不同学生或学生群体的测验结果具有可比性。同时,可减少无关因素对测验结果的影响,从而使之更为准确、可靠。

(3) 客观性

测验的标准化是为了提高测验的客观性,但绝对的标准化在现实教学中难以做到,因此,人们通常采用一系列指标来考察测验的客观性:①测验项目分析,通常所说的测验题目在教育心理测量学中被称为“项目”,项目分析即对测验题目的难度和区分度等进行分析,确保测验题目能合理地地区分出学习成功和不成功的学生;②信度,指测验结果的可靠程度;③效度,指测验结果的有效程度。

3. 考试

在教育测量领域,经常出现“考试”一词,比如中学毕业考试、大学入学考试,等等。考试与测验存在一些差异,主要有:

(1) 范围的不同

测验的范围较广,如心理测验、各种随机的小测验等。而考试一般是指比较正式场合下的测验,比如说期末考试、高考、自学考试等。要根据二者分别适用的场合合理使用两个术语,有些场合下两者可以相互替换,如既可说“单元测验”又可说“单元考试”;有些场合下,尤其是对一些公认的特殊称谓,两者不可相互替代,如“高考”不能说“高测”,“心理测验”不能称为“心理考试”等。

(2) 目标的稳定性

测验的目标一经确定,测验的内容只会在一定的范围内变化,而且,测验经过对某些群体的反复试用、修改后,会逐步形成标准化的测验;而考试目的性很强,

大多在某一特殊的时间和地域范围内为了某一特殊的目的而实施,因为下一次考试的目的和范围会发生很大的改变,所以难以在原有考试的基础上进行连续性的修改,因此,旨在测量学生对知识、技能掌握程度的考试(如高考)很难实现标准化;当然,以测量学生能力为主的考试也可实现标准化的过程,如 GRE(Graduate Record Examination)。

“考试是评价的主要方式之一,考试应与其他评价方式相结合,要根据考试的目的、性质、内容和对象,选择相应的考试方法。要充分利用考试促进每个学生的进步。”——《教育部关于积极推进高等院校评价与考试制度改革的通知》(教基[2002]26号)

4. 评价

评价是一个非常宽泛的范畴,包括了各种用来确定学生达到教学预期的学习目标的程度的方法,既包括测验又包括多种表现性评价方法。例如,为了评价一个学生处理多媒体信息的能力,可以使用客观性测验(objective test)调查学生掌握的多媒体工具方面的知识,再让学生在规定的实际操作任务中来搜集、处理多媒体信息并制作多媒体作品,来考察学生实际的多媒体信息处理技能。所谓多媒体只是测验中包含了大量多媒体知识的样本,其数量之大,远远超过了学生在实际操作任务中需要使用的知识技能。知识测验回答的是学生对要操作的技能“知道了多少”,表现性评价回答的则是学生“实际能做得如何”。本书主要关注信息技术教育过程中学生学习成就的评价(achievement assessment),重在通过各种各样的评价方法了解学生在信息技术教学干预下,对预期学习目标的掌握程度。为了便于陈述,在全书中都用“评价”来表示成就评价。

同时,尽管我们倡导尽可能多地使用真实性的评价,但由于受学校环境、时间、设备等方面的限制,往往需要将纸笔测验和表现性评价结合使用,才能做到完整地评价学生的信息技术学习成就。所以,本书中所用的评价是同时包含了纸笔测验和表现性评价的成就评价。

现在对测验之外的评价方式出现了很多种称谓,表 10-3-1 列出了与表现性评价有关的几个术语及其解释。

表 10-3-1

与表现性评价有关的常用术语

表现性评价 (performance assessment 或 performance-based assessment)	是指要求学生通过实际操作某项任务或一系列任务(如制作一个信息技术作品、利用信息技术开展一项研究等)来表现出他们的理解水平和操作技能水平的评价
另类评价 (alternative assessment)	是表现性评价的另外一种称谓,强调这些评价方法提供了有别于传统纸笔测验的其他一些评价方式
真实性的评价 (authentic assessment)	是表现性评价的另外一种称谓,强调在评价时,关注学生将理解和操作技能应用于真实世界中实际问题的重要性

二、参照点与单位

在利用测量原理进行量化的教育评价中都会用到参照点和单位这两个要素。

1. 参照点

参照点是计算的起点,参照点不同,测量结果就会因其所代表的意义不同而无法进行比较。理想的参照点是绝对零点。但是在教育和心理测量中,由于所要测量的特质常常很模糊,所以很难找到绝对零点,教育评价中的参照点一般随所参照的标准而变化。教育评价中主要采用两种参照标准对评价结果进行解释:常模参照(norm-referenced)和效标参照(criterion-referenced),采用这两种参照标准对评价结果做出解释分别被称为“常模参照解释”(norm-referenced interpretation)和“效标参照解释”(criterion-referenced interpretation)。

一次对学生学习成绩的评价可以提供:①学生的相对等级;②学生能够或不能够操作某项任务的描述。第一种结果表示的是学生在全部学生中所处的相对位置(例如,“在某个测验中,学生 A 在全班 35 个学生中排第三名”)。这种解释学

生绩效的方法称做常模参照解释;第二种结果表示的是对每个学生在具体知识和某些能力上掌握水平的描述,例如,“她能够分析一段程序的基本结构并描述它的用途”。这种解释评价结果的方法称做标准参照解释(与标准参照解释有关的术语见表 10-3-2。)

表 10-3-2

与标准参照解释有关的术语

领域参照解释	在一系列有关的、明确界定的任务范围(即领域)内解释评价结果。其含义与标准参照解释类似,这种解释描述性较强,较少使用
内容参照解释	与领域参照解释基本相同,但这里的内容领域更宽泛,即包括代表内容的任务也包括代表过程的任务。本术语也较少使用,多被标准参照解释代替
目标参照解释	评价结果是用某一系列试题所代表的具体目标来解释的。它也经常被称为标准参照解释,但主要用于局限于对某一非常具体的目标的解释

在具体教学中,这两种方法各有所长,常模参照评价能告诉我们学生与其他同学比较的结果,标准参照评价则能告诉我们学生对某项技能掌握的情况,而不必考虑其他学生的表现(参见表 10-3-3 对两者的比较)。大多数时候,这两种方法被视为统一的而非分离开的,在开展成就测验时往往是将两者的优点结合起来使用。

表 10-3-3

常模参照和标准参照评价的比较

	常模参照测验	标准参照测验
主要用途	调查性测验	掌握性测验
主要侧重点	测量学生在成绩上的个体差异	描述学生可以完成的任务
结果的解释	将学生的成绩与其他学生的结果进行比较	将学生的表现与一个明确界定的成就范围进行比较
内容的覆盖面	通常覆盖广泛的成就领域	通常限定在某些特定学习任务中
测验计划的特点	通常使用明细表	倾向使用详细具体的范围说明

	常模参照测验	标准参照测验
项目选择的程序	选择能最大程度将学生区分开的评价项目(例如,能得到一个可靠的等级),简单的评价项目通常会从测验中删去	包括所有可用于充分描述学生表现的所有评价项目。无须过分考量评价项目的难度或为了增加分数的范围而删去简单的项目
绩效标准	通过在某一特定群体中的相对位置来确定学生的绩效水平,例如,20个人中排第五名	通过绝对标准确定学生的绩效水平,例如,学生能说出90%的技术术语的定义,以此来描述学生的掌握水平

2. 单位

理想的单位应具备两个条件:一是要有确定的意义,即所有人对同一“单位”所代表的含义的理解相同,例如,所有人对“1千米”的理解都相同;二是单位要具有相同的价值,即相邻两个单位点间的差别相等,比如一千米和两千米之间以及二千米和三千米之间同样有一千米的差距。但教育测量中得到的数据的单位往往是不等距的或者是等距不等值的。例如:在一次信息技术考试中学生A得到了90分,学生B得到了80分,学生C得到70分,尽管90分与80分和80分与70分之间的原始分差相同,但却不等距,也就是说,不能据此判断学生A与学生B学习成绩的差异和学生B说学生C学习成绩的差距相同;同样,如果本次考试中,一道难题和一道相对容易的题目满分都是5分,这两个5分是不等值的。因此,分数等值问题是教育评价中非常重要但极易被忽视的问题,本书后面的章节将专门讨论转化原始分数以使其等值、等距的方法。

(三)、教育评价中的量表

教育评价中的量表(scale)的原意是“称”,也就是有零点并具有等距刻度,可用于给事物指派数字的测量工具。在教育评价中,量表可被定义为一个有单位和参照点,可根据测量目的和测量规划对被测事物赋值的连续体。根据量表的参照点和单位不同,从低级到高级、从模糊到精确将量表分成四类:

1. 命名量表

也称类别量表,是值根据某种法则给事物及属性指派名称或类别的量表,是最为简单、最为粗糙的量表,例如,在电子表格或数据库中常用的用数字进行编号,或用数字给不同的群体命名(用1代表男生,0代表女生)等等。命名量表数据或符号,只有区分性,但没有序列性、等距性、可加性等,不能进行代数运算和数量化分析,只适于进行次数统计,如次数、众数、百分比、离散相关等。

2. 顺序量表

也称等级量表,这种量表不仅将事物及属性分类,而且还在此基础上根据统一标准将之分为若干等级或进行顺序排列。因此,顺序量表比命名量表更为精确些,比如,根据学生的实际表现分别给予“优(5)”“良(4)”“中(3)”“合格(2)”“不合格(1)”等。这种量表所得的数据或符号具备了序列性,但仍没有等距性和可加性,仍不能对之进行加、减、乘、除运算。它所适用的统计有中位数、百分位数、等级相关系数、肯德尔和谐系数以及秩次方差分析等。

3. 等距量表

也称区间量表,它在给事物及属性指派数字和符号时,要求各数字或等级之间的差距是相同的,即具有相等的单位。本书后而讲到的标准分数即是一种等距量表。这种量表所得数字既具备了区分性和序列性,同时又具备了等距性和可加性。我们可加减或乘除同一个常数于每一个观测值,而并不破坏这些数值的关系,因此在一个等距量表上所得到的测值可以转换到另一个与该组资料计算的参照点和单位不同的等距量表上去。但是,由于这种量表的单位是人定零点,因此不能将其数据直接乘除,以倍数来解释。此种量表能最广泛地应用统计方法,如计算均数、变差、相关系数等统计量以及应用 t 、 F 检验等。

4. 比率量表

这是测量的最高水平,也是一种理想的量表。它有相等单位和绝对零点,因此所得结果不仅可以比较差距,而且还可以计算比例,以倍数来解释数据。很多物理测量都属于这一类量表,比如长度、重量、时间等测量。而教育和心理测量由

于难以确定绝对零点,因此很难达到这一量表水平。

(四)、评价的效度和信度

在设计评价计划时应充分考虑评价的效度和信度,即怎样才能够提供有效且可信的评价结果解释。效度指的是从评价结果中获取的信息的正确性和意义性,即评价所测的是不是评价者希望测量的内容。信度指的是评价结果的一致性。在后面的章节中会对效度和信度以及获得效度和信度的方法进行具体的讨论。

表 10-3-4 列出了提高在准备和实施评价时有利于提高评价效度和信度的建议。从表中可以看出,效度和信度是在拟定评价计划和准备评价的早期阶段就获得的,而不是在评价结果产生之后才有的。表中所列的程序将在“测验和操作基础上评价的准备”一章中详细讨论。这里仅仅是想强调一下它们在提高效率度和信度方面的重要性。

表 10-3-4

增加评价结果效度和信度的建议

理想的评价特征	应遵循的程序
1. 明确列出学习成果	1. 使用表现性的术语陈述学习目标
2. 从明确界定的学习任务的范围中选取代表性的样本	2. 准备一份关于计划评价的学习成绩范围的说明以及对所用评价任务样本的说明
3. 与计划测量学习结果相关的任务	3. 使评价任务与学习结果中所描述的具体操作相匹配
4. 评价任务的难度水平适宜	4. 使评价任务的难度与学习任务的要求、学生的能力以及评价结果的潜在应用相匹配
5. 能有效区别达到学习目标学生与未达到学习目标学生的任务	5. 遵循准备评价流程的一般指导原则和具体规则,避免可能歪曲普及结果的各种因素
6. 保证测量任务样本充分,以及得到的学生成绩样本充足;以提供可靠的结果并保证对结果的有意义的解释	6. 如果学生的年龄和评价时间限制了选用的任务数量,应对结果做出谨慎的解释,增加评价的频率,并使用其他证据验证结果
7. 采用的评价程序应有利于评价的准备和使用	7. 撰写明晰的指导语,评价程序的安排应便于施测、评分和解释

二、常用教育评价方法的分类

问题:

几乎每个教师在测验中都使用过填空题、选择题、匹配题、简答题、论述题,信息技术教师还经常使用以学生上机操作为基础的表现性评价,有些教师还尝试使用过更为综合的表现性评价方法。

你认为这些评价方法的各自的优点是什么?又各自存在哪些局限?你在实际的教学中有没有注意根据具体的评价需要选择最合适的评价方法?你在选用评价方法时,主要考虑哪些因素?

(一)、教育评价方法的分类

根据评价目标的不同要求,可将教育评价分为成就评价、能力测验和能力倾向测验三大类。如前所述,本书主要关注针对学生学习过程和学习结果的成就测验,所以对能力测验和能力倾向测验不做具体阐述。根据成就评价的内容,可将成就评价分为单科评价和综合评价两种,信息技术教育评价主要关注学生在信息技术学习中的成就评价,但应尽量体现课程综合化的需要;根据成就评价对应的教学阶段及其对教学和学习的功能,可将评价分为安置性评价、过程性评价和总结性评价,本书将在第二章结合评价与教学的关系详细叙述。

根据传统的分类方法,会根据评价采用的测验题目的特性,将测验分为客观性测验与主观性测验两大类。但随着评价理论的发展,人们日益认识到客观性测验和主观性测验并非是截然独立的两个极端,现实教学中采用的评价方法在评价任务的真实性、评价任务的复杂性、评价所需的时间、计分所需的判断能力等四个方面均处于一个渐变的连续体上(见表 10-3-5)根据这四个方面特性的区别,可将教学中常用的评价方法分为以下四大类型:

1. 选择—反应(selected-response)测验

要求学生从提供的各个答案中选择正确的或最佳的答案,包括选择题(multiple-choice)测验、是非题(true-false)测验和匹配题(matching)测验。

2. 补充—反应(supply-response)测验

要求学生通过补充一个单词、短语或一篇完整的论文做出反应,主要包括填空题、简答题等。

3. 限定性表现性(restricted performance)评价

关注学生在一个结构良好的限制性的任务中的实际表现。例如,按照明确的要求给一段文本设置格式;按要求制作一个多媒体作品;按要求写出一段程序语言等;或者根据提供的关键词使用搜索引擎查找信息等。

4. 拓展性的表现性(extended performance)评价

这种评价则涉及更综合、结构化较差的操作性任务,例如,根据对一个名人的研究用文字处理软件写一篇人物传记;根据对当地生活污水的处理状况,制作一篇关于环境保护的多媒体演示文稿;或者使用计算机软件分析当地近百年来的气候变化等。除了更多地使用拓展性的表现性评价,现在的信息技术评价一般都要学生在现实情境中,有机结合和应用学过的知识、技能去完成任务。如果该任务要求学生完成信息技术作品的话,一般还要求学生提交产品之前对作品进行的评价和修改。

表 10-3-5 四种评价方法的比较总结

测验			表现性评价	
选择-反应 测验	补充-反应 测验		限定性 表现性评价	拓展性 表现性评价
任务的真实性				
低	----->			高
任务的复杂性				
低	----->			高
评价所需时间				
低	----->			高
计分所需的判断能力				
低	----->			高

在现实教学中,教师偏爱选择型测验(包括选择题、是非题和匹配题)的出发

点是可以在相对较短的时间内完成大量的问题,而且便于施测和评分,测验结果又以数字的形式呈现,易于记录、比较及向他人报告。但教师使用的选择型测验的内容过分局限在对事实性知识和名词的记忆上,大量研究(Gronlund N. E, 2003)表明 80%~90%的教师自编测验都是以知识类的学习成果作为考察对象。正是这种过分强调简单知识的机械记忆和简单复现为主要特征的选择型测验招致了大量的批评,教育界开始广泛重视在实际情境中考察复杂的学习结果,并由此引发了评价技术改革的运动。但在追求表现性评价的这场运动中,人们往往忽视了这样一个事实——只要设计合理,纸笔测验也能考察广泛的复杂的学习结果。在现实的教学中,只有同时使用客观性测验和表现性评价的教学才能取得理想的效果。

(二)常用教育评价方法的分析

以下是对各种常用评价方法从评价任务的真实性、评价任务的复杂性、评价所需的时间、计分所需的判断能力等四个方面进行的比较:

1. 评价任务的真实性

任务的真实性是指评价任务在多大程度上模拟了真实生活中的表现。传统的选择型测验真实性很低,因为它要求被评价者从给出的一系列可能答案中选择一个作为答案,学生的反应被限定在罗列出的选项中,而这种高度结构化的问题在真实生活中很少发生。拓展性的表现性评价具有很高的真实性,因为它试图完全模拟真实生活中的表现。当我们评价学生某种软件操作的熟练程度如何,作品设计和制作得好不好,或者理解现实生活中信息问题的能力怎样(例如,如何在最短的时间内搜索到最有针对性和最有价值的信息等)时,要求学生做出一系列与在真实生活中几乎一样的反应。处于这两个极端之间的是补充—反应测验(例如填空题、简答题和论述题等)和限定性表现性评价,它们所给任务的结构化程度中等,所要求的反应较自由,因此比选择—反应测验更具有真实性。

在人们日益提倡拓展性的表现性评价以增加评价的真实性的同时,还呈现出

另一种趋势,即增强传统纸笔测验的仿真性,测验中使用的问题和测验的过程更接近于真实生活,使纸笔测验也可以考察更复杂的学习结果。以一道关于“算法”的纸笔测验题为例,题目中呈现的信息可能会远远超过学生解决一个所需的信息,藉此考察从中选择和处理有用信息并整理解决问题思路的能力。另外,有些测验从选择型向补充型转变,或者采用综合运用两种类型的测验,例如,在学生选择一个选项后,要求学生简要解释选择该答案的理由等。

2. 评价任务的复杂性

选择—反应类的测验试题所呈现的问题和期望学生所做出的反应都不复杂。尽管这类试题也可以考察学生的理解能力和思维能力,但它们的典型方式是呈现一个单一的、有明确限定条件的问题,并要求学生选出惟一正确或最佳的答案。而拓展性的表现性评价则涉及多种学习结果,需要有机结合来自不同来源的观点和技能解决问题,需要采取各种可能的解决方案,并需要对结果从多维度进行评价。例如,综合考虑各种因素设计选择最佳旅游计划的算法等。就任务的复杂性而言,补充型测验也处于两个极端之间。以论述题为例,它可以考察学生选择、整合和表述观点的能力,但短文写作的任务与表现性评价的任务相比,限定性更强,结构化程度更高。

3. 评价所需的时间

选择型试题可以在一个相对较短的时间内让学生完成大量的任务,评价结果可以依靠人工或机器迅速评出分数。这种测验的高效率无疑是其能够被广泛使用的一个主要原因。表现性评价所需的评价时间则是最长的。这类评价中的一些任务可能需要几天甚至几周去完成,例如,研究性学习任务。还有一些表现性任务可能一次只能考察一个或少数学生,例如,计算机操作、研究结果演讲等。通常评定这些表现性任务的过程和结果也比较费时、费力。补充型测验,比如论述题所需的时间要远远多于选择型测验但少于表现性评价。

由于表现性评价花费的时间较多,能够纳入教学计划的评价问题数量有限,

会导致评价任务较少,使考察的内容不全面。由此引起一个核心的评价问题,即评价的结果或学生在一个表现性评价任务中的表现在多大程度上能被推延到与之相似的任务中去。表现性评价可以给学生呈现“真实生活”中的问题,但问题通常是某个具体情境中的独一无二的问题,而现实生活又是不断变化的。所以,学习的迁移是表现性评价中要考虑的关键问题。在现实的教学过程中,需要慎重考虑在表现性评价上花费的时间是否必要,一般来说,只有当评价是教学过程中的有机组成部分,而且教学重视学生学习结果的可迁移性时,例如,需要培养学生的推理能力、批判性思维能力等,才有必要花费大量时间开展表现性评价。

4. 计分所需的判断能力

不同的评价方法需要做出的判断相差很大。选择型试题的反应有明确的对错之分,由此得到的分数具有完全的客观性,即不同的评分者会对同一反应给出同样的分数。论述题允许学生有较为自由的反应,这就使得最后的分数带有较强的主观性色彩,不同的评分者会对答案的不同要素(例如,短文的完整性、短文的组织结构、书写的清晰程度等)赋予不同的权重,个人的偏见也会对评分者造成一定的影响,所以不同评分者会对同一反应给出不同的分数。任务越综合、越复杂,比如在表现性评价中,对教师评分的要求就越高。复杂的表现性任务是各种类型信息和学生技能的结合,可能存在多种解决途径。很多情况下,表现性任务难以也不应有一个像在短文写作中常用的那种示范性答案。鉴于此,表现性评价必须依靠关于优秀表现的标准,然后依据这套标准制定等级量表或计分量规。但每一步都建立在主观判断的基础之上。

上述对四种类型评价的评估揭示了每种方法都有各自的优点和缺点。选择一反应测验能涵盖全而的课程教学内容,便于施测、评分和解释评价结果,但牺牲了真实性和某些复杂性;拓展性的表现性评价可以保证评价具有高度的真实性,提高评价的复杂性,但往往会耗费过多的时间,对学生表现的评估也需要更多的主观判断。在处理这个矛盾上行之有效的原则是:只要评价方法适于计划考察的

学习结果,就尽量选择效率最高的评价方法,同时,不要因为对复杂学习结果的评价太耗时、结果难以评定就忽略对复杂学习结果的评价。

第二节 有利于促进学生学习的信息技术教学评价观

一、评价与教学的关系

(一)设计合理、实施恰当的评价是有效教学的有机组成部分

在传统的教学设计中,人们往往只在教学结束后实施评价,更多地将评价作为测量学生学习结果并分析学生是否掌握有关知识技能的工具,而忽视评价与教学过程和学习过程不可分割的内在联系。与此相联系,在教学中大多独立进行教学设计和评价设计,难以真正落实评价对教学的促进作用和对学习的诊断作用。近年来,教育评价界日益认识到在教学活动的各个阶级,评价均能通过多种方针提高教学质量并促进学生的学习,一个设计良好、实施合理的评价过程应是一个有效教学过程密不可分的组成部分并应与教学过程平行展开,一个有效评价过程与有效教学在特征上高度吻合(如表 10-3-6 所示)。因此,信息技术教师应逐步树立“教学—学习—评价”有机结合的教学评价观,将对教学评价的设计和规划融合到教学设计中,使其成为教学设计中不可或缺的组成部分,将仅注重“教—学”过程的教学设计和规划转变为对“教—学—评价”过程的全面设计,确保评价对教学 and 学习的“全过程”的促进作用。

表 10-3-6

教学与评价之间的关系

教学 要取得最有效的教学效果,须做到:	评价 要取得最有效的教学效果,须做到:
1. 教学指向一系列界定明确的预期学习结果	1. 评价是被设计来考察一系列界定明确的预期学习结果
2. 教学方法和材料适合计划达到的教学目标	2. 评价的性质和功能适合于要评价的目标
3. 教学的设计符合学生的特点和需要	3. 评价的设计符合学生的特点且对所有学生都是公平的
4. 教学决策是以有意义的、可靠的、相关的信息为基础做出的	4. 评价提供的信息应当是有意义的、可靠的、相关的
5. 定期告知学生他们的学习进展状况	5. 尽早给学生提供关于评价结果的反馈
6. 对于没有达到要求的学生给予及时的补救	6. 通过评价结果能够揭示具体的学习不足
7. 定期评估教学效果,根据需要调整预期学习目标和教学方法	7. 评价结果能为评价教学目标、教学方法和教学材料的适合性提供有用的信息

(二)通过全过程评价发挥评价对教学与学习的诊断、激励和促进作用:理念与方法

为了与教学保持完全一致,对评价的规划应当与对教学的计划同步进行,从教学前、教学过程中到教学结束的整个过程中教师需要做大量的决策,在此过程中,精心规划的评价程序能够为教师提供更加客观、全面的信息,帮助教师对教学状况进行判断并做出更为有效的决策。

1. 教学前的评价——安置性评价

建构主义学习观认为,学习是学习者利用已有知识经验了解建构新知识的过程,信息技术的学习也不例外。因此,规范的信息技术教学设计要求教师在教学之前对学习者的已有知识经验进行分析。在分析学习者的知识经验时,教师往往需要考虑两个问题:①对开展新的教学所必需的知识、技能,学生掌握到了何种程

度? ②对计划进行的教学的预期学习结果,学生已经具备到何种程度? 教师可利用自己的教学经验对这两个问题做出回答,也可以通过各种形式的安置性评价来获取关于学习者已有知识基础的更全面的信息。对于不熟悉学生情况的新教师而言,下述用以考察学生知识技能基础的测验尤为重要。

(1)常用的安置性评价

①准备状态前测

关于第一个问题,可通过准备状态前测(readiness pretest)来获得有关信息,这种测验一般在一门课程或一个单元的教学开始前进行,用来检验学生是否具备了学习新课程所必须的知识和技能。例如,在学习高中信息技术的某一选修模块之前,可以先通过与该选修模块对应的必修模块测试考察学生的准备状态,如果测验显示学生缺乏学习该选修模块必备的知识技能,教师就应有针对性地进行补救,或者根据学生现有的水平调整教学内容和教学难度。

②安置性前测

关于第二个问题,可通过安置性前测(placement pretest)来回答。安置性前测与在教学活动结束后进行的测验几乎相同,即安置性前测采用的是教学活动结束后所用测验的副本。安置性前测可了解学生是否在教学之前就已经掌握了在教学计划中安排的内容——如果学生已经掌握,教师就需要调整后续教学计划,例如可鼓励学生跳过某些教学内容或组织学生开展拓展性的学习;如果学生尚未掌握,则应循序渐进地学习。图 10-3-1 总结了安置性评价的作用。

③利用表现性评价确定学生起点水平

除了使用测验外,还可以采用表现性任务来确定学生的起点水平。例如,在进行高中“信息技术基础”中“信息加工与表达”部分的教学之前,通过一个需要学生完成的对文本、多媒体等信息进行加工处理的综合任务,获得学生在这方面的实际表现水平,为随后的教学及其评价提供参考依据。同时,如果教师计划使